Research Data Management Implementation at Peking University Library: Foster and Promote Open Science and Open Data

Hua Nie[1†], Pengcheng Luo[1] & Ping Fu[2]

[1]Peking University Library, Beijing 100871, China

[2]Central Washington University Library, 400 E. University Way, Ellensburg WA 98926, USA

ABSTRACT

Research Data Management (RDM) has become increasingly important for more and more academic institutions. Using the Peking University Open Research Data Repository (PKU-ORDR) project as an example, this paper will review a library-based university-wide open research data repository project and related RDM services implementation process including project kickoff, needs assessment, partnerships establishment, software investigation and selection, software customization, as well as data curation services and training. Through the review, some issues revealed during the stages of the implementation process are also discussed and addressed in the paper such as awareness of research data, demands from data providers and users, data policies and requirements from home institution, requirements from funding agencies and publishers, the collaboration between administrative units and libraries, and concerns from data providers and users. The significance of the study is that the paper shows an example of creating an Open Data repository and RDM services for other Chinese academic libraries planning to implement their RDM services for their home institutions. The authors of the paper have also observed since the PKU-ORDR and RDM services implemented in 2015, the Peking University Library (PKUL) has helped numerous researchers to support the entire research life cycle and enhanced Open Science (OS) practices on campus, as well as impacted the national OS movement in China through various national events and activities hosted by the PKUL.

---

† Corresponding author: Hua Nie (Email: hnie@lib.pku.edu.cn; ORCID: 0000-0002-4522-5049).

1. INTRODUCTION

Open Science (OS) has made science more efficient, reliable, and responsive to societal challenges and reshaped scholarly communication landscapes [1]. Open access to research data is regarded as a driver for OS [2]. The scientific research community believes that data should be open, accessible, and reusable. Data sharing and reuse help strengthen researchers' and institutional data stewardship [3,4,5]. To better foster OS, Open Data (OD), and transformation of scholarly communication, more and more academic libraries have provided or plan to provide library-based research data management (RDM) services for their home institutions. Tenopir et al.'s study [6] suggested that academic libraries be ideal centers for research data service activities on campuses, providing unique opportunities for academic libraries to become even more active participants in the knowledge creation cycle in their institutions. More recent studies agreed that academic libraries and academic librarians, as active stakeholders, have been playing a significant role in fostering open access movement, transforming scholarly communication landscapes, and facilitating RDM services [6,7,8,9,10,11,12].

Although RDM has been rapidly developed in research universities in the world, RDM in Chinese academic libraries is still at the early stage of the development. For example, the re3data.org (the Registry of Research Data Repositories) is listed more than 2,000 research data repositories in 2019, while among them, there were only 42 research data repositories related to China. The re3data.org an OS tool that offers researchers, funding agencies, libraries, and publishers an overview of existing international repositories for research data [13].

Peking University Library (PKUL), as a top research university library in China, has long been actively seeking opportunities to raise awareness, foster collaboration, and initiate projects from both inside and outside of the University. The PKUL also has been actively involved in numerous national and international endeavors to foster open access movement and transform scholarly communication for the past decade. Since 2010, the PKUL has implemented the following initiatives to support the dynamic changing environment of the scholarly communication in the University: Peking University (PKU) Institutional Repository in 2010, PKU Open Journals in 2015, Scholars @ PKU in 2015, and PKU Open Research Data Repository (PKU-ORDR) in 2015. Particularly, the PKU-ORDR was created in 2015 for facilitating more effective and efficient data preservation, data sharing and reuse, and providing incentives for making data readily accessible to researchers and for the general public.

This paper will review how the PKUL implemented the PKU-ORDR project and RDM services to foster and support OS and OD on campus and impact Chinese OS communities. The implementation phases being reviewed include project kickoff and needs assessment, partnerships establishment, software investigation and selection, software localization, and customization, as well as the implementation of RDM policies and services. Some issues revealed during the stages of the implementation process will be also discussed and addressed in the paper such as awareness of research data, demand from data providers and

users, data policies and requirements from home institution, funding agencies, and publishers, the collaboration between administrative units and libraries, and concerns from data providers and users.

The significance of the study is that PKU-ORDR shows a successful example of creating an OD repository and RDM services for other Chinese academic libraries planning to implement their RDM services for their home institutions in the future. The authors of the paper have also observed since the PKU-ORDR and RDM services implemented in 2015, the PKUL has helped numerous researchers to support the entire research life cycle and enhanced OS practices on campus, as well as impacted the national OS movement in China through hosting various national events and activities.

2. RELATED WORKS

The related works will focus on some aspects that this paper will address such as RDM and academic libraries & librarians, RDM and open research data repositories and systems, collaborations between research units and libraries, service support and promotions, repository implementation, data curation, and research support staff's or librarians' skills training.

Tenopir et al. [6] pointed out that science becomes more collaborative, data-intensive, and computational, and academic researchers face a series of data management needs. Meanwhile, Moon's study [14] shows that research funding agencies require researchers to provide DMPs when they apply for a grant and publishers also require researchers to provide data when publishing research results. Curdt's study [15] indicated that science conducted in cross-institutional, interdisciplinary, and long-term research projects requires active sharing of data, documents, and further information. Thus, RDM services should be established to support all researchers during their entire individual research studies.

Tenopir et al. [6] also claimed that academic libraries may be ideal centers for RDM service activities on campuses. Cox et al. [10] reported an international study of RDM activities, services, and capabilities in higher education libraries. Their study found that libraries have provided leadership in RDM, particularly in advocacy and policy development. However, services provided by libraries are still limited, focused especially on advisory and consultancy services. Tripathi et al. [16] studied the RDM services implemented by different university libraries in India for managing, organizing, curating, and preserving research data generated at their universities' departments and laboratories for data reuse and sharing and suggested a model for the university libraries to follow for actually deploying RDM services.

Johnston et al. [17] compared six institutions' RDM support levels within the Data Curation Network project and developed a shared staffing model for data curation across multiple institutions to support their researchers to meet their data-sharing goals through library-based data repository and curation services. Lee et al. [18] interviewed some American university institutional repositories (IRs) staff and then provided a rich, qualitative description of research data curation and use practices in IR. In particular, Lee et al.

identified data curation and use activities in IRs, as well as IRs structures, roles played, skills needed, contradictions and problems exposed, solutions sought, and workarounds applied.

Curdt and Hoffmeister [19] shared their design and implementation of RDM services for a multidisciplinary and collaborative research project. McKinney et al. [20] described that Harvard University established a diffraction data publication system, the Structural Biology Data Grid (SBDG[①]), to preserve primary experimental data sets supporting scientific publications. All data sets published through the SBDG are freely available to the research community under a public domain dedication license, with metadata compliant with the DataCite Schema[②]. They also shared their practices that the SBDG collaborated with the Institute for Quantitative Social Science at Harvard University to extend the Dataverse[③] open-source data repository system to structural biology data sets.

Mannheimer et al. [21] described how data repositories and academic libraries can partner with researchers to deal with challenges associated with qualitative data sharing and suggested that data repositories and academic libraries could help researchers address some of the challenges associated with ethical and lawful qualitative data sharing. Dovidonytė's study [22] described the Lithuanian landscape of OS policies and institutional involvement in OS practices. The author also discussed prerequisites for sustainable and consistent OS implementation such as OS infrastructure, incentives for researchers, research assessment, and repositories' compliance with the European Council requirements on a national level.

Pontika [23] made an analysis and found that academic libraries have created some new academic librarians' positions to support OS, OD, scholarly communication, and RDM on their campuses. However, researchers are still unfamiliar with RDM best practices, and research support staff including librarians is faced with the difficulty of providing support to researchers across different disciplines and career stages [24].

Alonso-Arévalo [25] agreed that the management of research data is one of the major challenges facing scientific and research libraries in the coming years. Already half of the American universities have a work plan on this issue, and all trend reports agree that RDM will be one of the priorities and future issues to be taken up by research libraries. Söderholm et al. [26] found that the network-based collaboration model that fosters individuals' interconnectedness is crucial for surviving with the built-in dynamism of RDM. Tang and Hu emphasized in their study [27] that for growing RDM services, institutional commitment to resources and training opportunities is crucial. As an emergent profession, data librarians need to be nurtured, mentored, and further trained.

All of these studies have provided some theoretical, useful, and practical insights and examples for us and also showed us some challenges and issues in the RDM implementation process faced by researchers, academic libraries, and librarians.

[①] data.sbgrid.org

[②] schema.datacite.org

[③] dataverse.org

3. RDM IMPLEMENTATION

3.1 Kick-Off and Needs Assessment

As the kicking off of the PKU-ORDR project, the PKUL conducted a campus-wide survey to get a better understanding of RDM needs and requirements from researchers and research teams in 2013. The purpose of the survey was to identify the real needs of researchers and collect data from them so that the PKUL can create a strategic roadmap and steps to create a framework or a platform to meet the needs of RDM. The analysis results were summarized and published in the *Journal of Library and Information Service* [28]. The survey focused on the following aspects: awareness and current practices of RDM including data preservation, sharing, and reuse; description and features of research data; the current state of RDM; and expectations of RDM services. The survey results showed that 87.5% of respondents were willing to share research data under certain conditions. The biggest motivation that they were willing to share was because the participants recognized the value of sharing data, the positive relation between data use and citations, data visibilities, and credits awarded to data providers. However, the biggest concern for researchers was the issue of plagiarism.

The PKUL also interviewed 23 research teams from multiple disciplines on the campus. The face to face communication with the teams helped discover more valuable information about the current state of RDM including long term preservation, data sharing, and data reuse. Zhu et al. [28] summarized three major findings from the interviews: (1) Research data sharing behavior is significantly influenced by disciplines. For example, biology is a data-driven and data-intensive discipline in which open access has already been a common best practice, data sharing standards and norms were already well established and put in place. (2) An embargo period with data sharing is generally expected and required. Almost all researchers being interviewed emphasized that their data should be shared after their results are formally published, which addressed the concern of possible plagiarism. (3) Data sharing behavior is more spontaneous and passive than active and lacks proper incentives and necessary maintenance, as well as a well-established mechanism for data citation, recognition and credits, and feedback from data users.

The interview also revealed that the data management and needs of researchers in different disciplines vary greatly. Bioinformatics researchers need very large data storage so that the large amounts of process data generated from their experiments can be preserved. Researchers from Computer Science are willing to share their data; however, Computer Science data are often considered very large data and make data sharing cost more expensive. For example, the volume of the Chinese Web data set collected by the Institute of Network Computing and Information Systems in the past ten years is above 100TB. Researchers from Business hope they can obtain more valuable enterprise and government data that can be used in their classes and research. Researchers from the Institute of Social Science Survey (ISSS) hope to maximize the value of their survey data as much as possible through data sharing; however, the ISSS established a relatively strict user application procedure for users to access data. Faced with so many different data management needs of researchers, as an initial attempt, the PKUL analyzed the data needs based on priorities and decided to build an initial service infrastructure to meet the needs with the highest priorities. Due to a variety of process data associated with different disciplines, the PKUL decided to focus on data

closer to the final state and easier to share and collaborate with institutions inside and outside the campus to build the PKU-ORDR, making data easier for PKU researchers to access, reuse and share.

3.2 The Establishment of the Collaborative Model

As one of the most active advocates of RDM at the University, the PUKL made numerous efforts to convince various university research administrative units to invest in and provide support to create a library-based RDM framework. The PKUL also sought some potential partners within the University since cooperation and collaborations with administrative units and other units on campus are vital to the success of the project and critical for the sustainability of the project.

The PKUL finally selected the Institute of Social Science Survey (ISSS) as a working partner to cooperate and collaborate on the development of the RDM services. The ISSS was created to act as a social science data survey coordinator and interdisciplinary empirical research platform that enables Peking University as well as other research institutions around the world to study China's social problems and conduct social science research, mainly through undertaking large-scale social survey projects and sharing the survey data openly. So the ISSS was an ideal collaborative candidate for the Library. The ISSS also plays a leading role on campus to provide workshops and training classes in data access, curation, and methods of analysis for the social science research community.

In 2014, the Peking University was awarded a grant by the National Natural Science Foundation of China for the China Survey Data Archive (CSDA) project, which aimed to develop a data repository administrated by the University Management Science Data Center (MSDC), a department within the ISSS. This grant provided an opportunity for the PKUL to build a more collaborative relationship with the ISSS. With the assistance of the research administrative units such as the Office of Science Research and the Office of Social Science Research, the PKUL and the ISSS decided to work together on this project. Initially, the responsibilities were split as follows: The MSDC supervised by the ISSS was responsible for research data collection and cleaning-up, standardization and analysis, data repository platform testing, and feedback. The PKUL was responsible for requirements analysis, functional design, software selection, as well as the development and maintenance of data repository, data storage, classification and metadata, systems administration, and associated technical and technological services.

However, the ISSS and the PKUL soon discovered through the analysis of the data collected from the survey and interviews that it could be an opportunity to build a strong showcase for OD for the nation because there were only a very limited number of subject-specific and/or research team-oriented data storages and data services available either at the institutional level or at the national level. The initiative was named as PKU-ORDR (PKU Open Research Data Repository) project as a sub-project of the CSDA project, with its goal to develop an infrastructure to support PKU researchers to manage their data more effectively and efficiently and provide RDM services ranging from storages to consultations.

The strategic objectives of the PKU-ORDR are summarized as below:

- To publish high-quality research data and disseminate academic outputs through an open platform;
- To promote OS, facilitate data sharing and reuse, and encourage to reproduce research;
- To enable and track data citations and usage metrics;
- To explore data publishing and long-term preservation solutions;
- To foster innovation and cross-disciplinary integration.

In addition to the ISSS, the PKUL also cooperated with other internal units and external organizations to enrich the data content of the PKU-ORDR. Through collaboration with the Center for Bioinformatics of Peking University, the PKUL created linked data in the PKU-ORDR linking to the Bioinformatics database. Through cooperation with the Beijing Information Resources Management Center, the PKU-ORDR interoperated with the Beijing Government Data Resource System (BGDRS) so that the registered users in the PKU-ORDR can download data from the BGDRS directly. Through cooperation with the National Information Center, the PKU-ORDR collected some valuable enterprise data sets across the country. All these collaborations have greatly enriched the data content and expanded the disciplines' scope of the PKU-ORDR.

3.3 The Establishment of the Open Research Data Repository

The first step of the PKU-ORDR was to create an open research data repository to meet the needs of data storage and data sharing. The establishment of the data repository includes selecting software as a framework and customizing the software.

3.3.1 Software Selection

There were some types of RDM software available at that time, including various institutional repositories (IRs) to support RDM, either open source solutions or proprietary solutions. The PKUL evaluated and assessed various types of existing software including Dataverse, Data Conservancy, CKAN, Dryad, ICPSR, Genbank, Figshare, and Nessta. The implementation team also deployed and tested some open-source solutions such as Dataverse, Data conservancy, CKAN, and DSpace.

The implementation team adopted a software metrics tool and created some criteria to evaluate and assess these software solutions. Some general criteria were considered such as business and industry expertise, market knowledge, program/project management capabilities, methodology, communications, and independence and objectivity. Besides, as shown in Table 1, four specific criteria were particularly considered: ① Metadata standard and interoperability; ② Permissions management and access control; ③ DOI identifier and version management; ④ Online analysis and visualization. It is noted that the Dataverse metadata schema consists of a compulsive citation metadata block and multiple optional discipline metadata blocks that can be easily customized. The default discipline metadata block is DDI for Social Sciences and the Dataverse also provides several other disciplines metadata blocks, such as Biomedical, Geospatial, Astronomy, and Astrophysics. Therefore, the Dataverse metadata schema is flexible enough and can adapt to any discipline theoretically.

<div align="center">

Table 1. Software comparisons.

</div>

Software	Type	Domain	Four specific criteria
Dataverse	Open source software	Multidisciplinary, mainly Social Sciences	Supporting ① based on DDI; supporting ②; supporting ③; supporting ④ based on TwoRavens
Data conservancy	Open source software	Multidisciplinary	Beta version: supporting ①; partially supporting ②; not supporting ③ or ④
CKAN	Open source software	Multidisciplinary, mainly government data	Partially supporting ②; not supporting ①, ③, and ④
Dryad	Open source software	Multidisciplinary, mainly Bioscience	Supporting ① based on expanded DC standard; not supporting ② or ④; partially supporting ③
ICPSR	Proprietary software	Social Sciences	Supporting ① based on DDI; supporting ②, ③, and ④
Genbank	Proprietary software	Bioinformatics	Supporting ① and ④; partially supporting ②; not supporting ③
Figshare	Commercial software	Multidisciplinary	Supporting ③; not supporting ①, ②, or ④
Nesstar	Commercial software	Social Sciences	Supporting ①, ②, and ④; not supporting ③

Note: ① Metadata standard and interoperability; ② Permissions management and access control; ③ DOI identifier and version management; ④ Online analysis and visualization.

After systematic comparisons and assessment, the Dataverse solution was finally chosen as the development tool. The Dataverse was originally developed by Harvard's Institute for Quantitative Social Science (IQSS), along with many collaborators and contributors worldwide. As of August 7, 2020, it has had 59 installations in the world.

3.3.2 Software Customization

Although the Dataverse was chosen as the framework for the open research data repository, customization was a challenge. The development of the software and version release phases is shown in Figure 1. The project milestones are summarized as below: (1) by the end of May 2015, the PKUL completed testing and functions building with the Chinese version of Dataverse v3.3 adopted from Fudan University. (2) Starting from June 2015, the PKUL continued working on system architecture, localization and customization, and functionality and features refinement based on Dataverse 4.0 which then was released and the PKU-ORDR was formally launched in December 2015. Version 4.0 was used between June 2015 and July 2019. During that time, Harvard University released more than ten minor versions with multiple functions added to the framework such as metadata harvesting, private URLs, and cloud storage support. (3) The PKUL decided to upgrade the platform to v4.14 by adopting changes made by Harvard University and fixing numerous bugs in the early v4.0 in July 2019. The PKUL also completed the function customization and data migration of the platform by the upgrade.

Figure 1. Version development of PKU data repository.

Here are the highlights of our local customization: (1) user management, (2) bilingual interface, (3) usage statistics, (4) data contests, (5) other functions such as DataCite DOI registration, data set related publications, and (6) custom home page.

To enhance user management function, the PKUL implemented the PKU-IAAA single sign-on system in the platform to enable our users to quickly and securely authenticate their permissions and instant access to the OD repository, and the relevant patron information can also be carried into the data repository through the PKU-IAAA. Furthermore, the PKUL enabled group download function so users can download multiple files within one data set with one request while original Dataverse only allows users to download one file with one request. Also, the PKUL created two types of user account: regular user account and advanced user account. A regular user account can be upgraded to an advanced user account when a user submits his/her application and provides more required information to get more privileges to become an advanced user. A bilingual interface is essential for our users since our repository is open to anyone in the world. Original Dataverse provides only unilingual descriptions. Researchers always publish their research outputs in English to increase the visibility of their research. From this perspective, the English language is an ideal candidate for the user interface. However, the majority of our users come from China, and the Chinese language is their mother language and more comfortable for them to use. So the PKUL decided to make the Dataverse repository interface and metadata support both English and Chinese. The user interface can be switched between Chinese and English and search results can be displayed both in English and in Chinese. The notifications sent to users are also be customized in a bilingual format.

Regarding usage statistics, the original Dataverse system only tracks the number of downloads, which is far from satisfying the needs of the PKU data providers' statistical requirements. Therefore, the PKUL enabled log records of the user application, administrator verification, user browsing, and download, etc. ElasticSearch is used to index the logs so that the data provider can query and download real-time data. Meanwhile, Baidu Analytics was implemented in the data repository pages to analyze data such as user sources, devices used, keywords for search, and pages visited. Furthermore, the PKUL hosted two national data contests to promote open research data repositories. The contest module was added to the Dataverse repository to facilitate user enrollment and data use. Participants were allowed to form teams to enroll in the contest, submit their papers, and access research data directly by using their user accounts of the data repository. The contest module also provided functions such as the contest homepage and paper display gallery.

Additionally, the PKUL added many other functions to Dataverse. Dataverse 4.0 only provided Handle identifier registration, and the PKU-ORDR adopted DataCite DOI to register data. Our module was later

adopted by Harvard University and other institutions that are using Dataverse. Since some data sets within the PKU-ORDR are of high quality, for example, the China Family Panel Studies data set has been cited by numerous research papers, the PKUL also used API to interoperate with the PKU-IR to retrieve papers from the PKU-IR and display those papers associated with the data sets on the PKU-ORDR platform. Also, Dataverse 4.0 did not support homepage customization but the PKUL developed a custom homepage and now such a homepage customization technology has been adopted by Harvard University. As shown in Figure 2, numerous efforts had been made between 2014 and 2019, and key milestones are highlighted in the diagram.
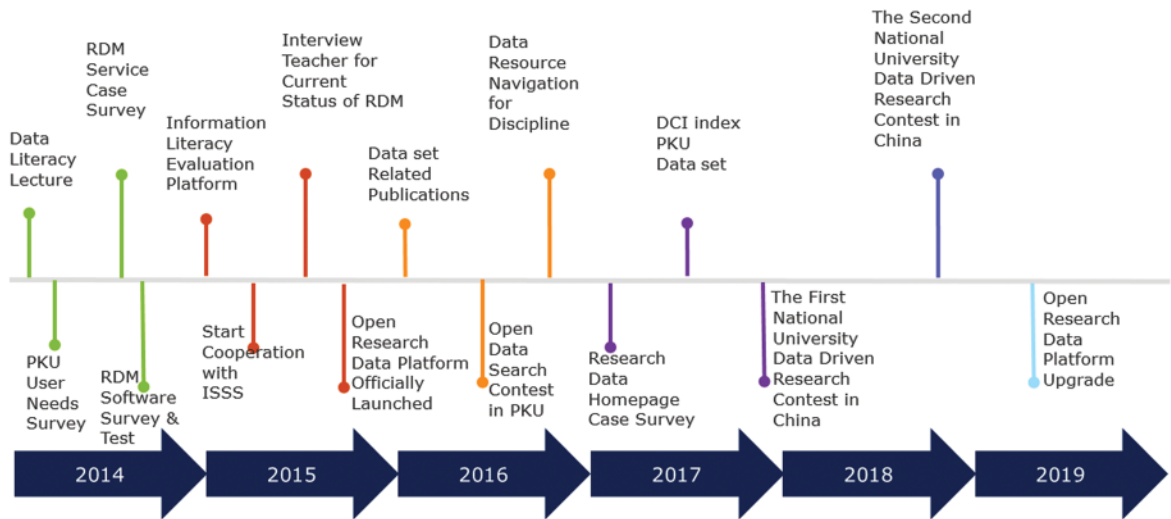
Figure 2. Research Data Management (RDM) service framework development diagram.

3.4 Usage, Data Curation, Skills Training, and Services Promotion

3.4.1 Usage

The PKU-ORDR enhanced the PKUL's infrastructure for data storage and sharing. Through collaborating with academic departments on campus, the PKU-ORDR has collected numerous high-quality data sets, examples include China Family Panel Studies, China Health and Retirement Longitudinal Study, and Beijing Area Study, Comprehensive Language Knowledge Base, and AutismKB, an Evidence-based Knowledge Base of Autism.

As of August of 2020, the PKU-ORDR has released 66 Dataverses, 305 data sets, and 2,036 data files. The total number of downloads has exceeded more than 620,000, The average number of daily visitors is about 500, and the average number of page views is 2,700. In recent years, there have been numerous visitors from more than 89 countries who visited the repository, and the top five countries are China, the United States, the United Kingdom, Japan, and South Korea. The number of registered users has reached 32,000. Figures 3 and 4 show respectively the top 10 institutions in terms of the registered users in China

and abroad. As shown in the figures, all these registered users came from prestigious research universities either in China or in the other part of the world.
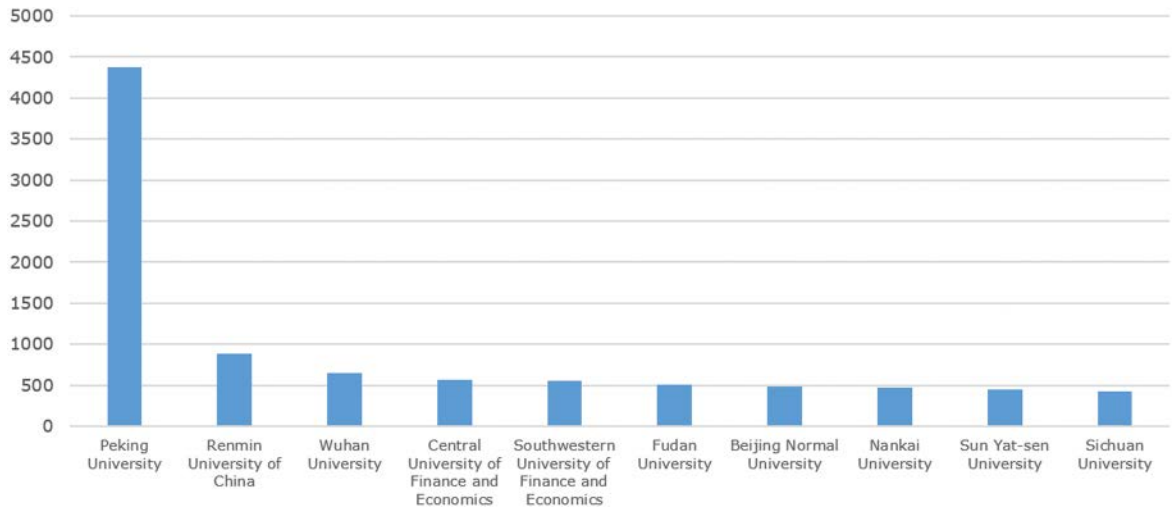

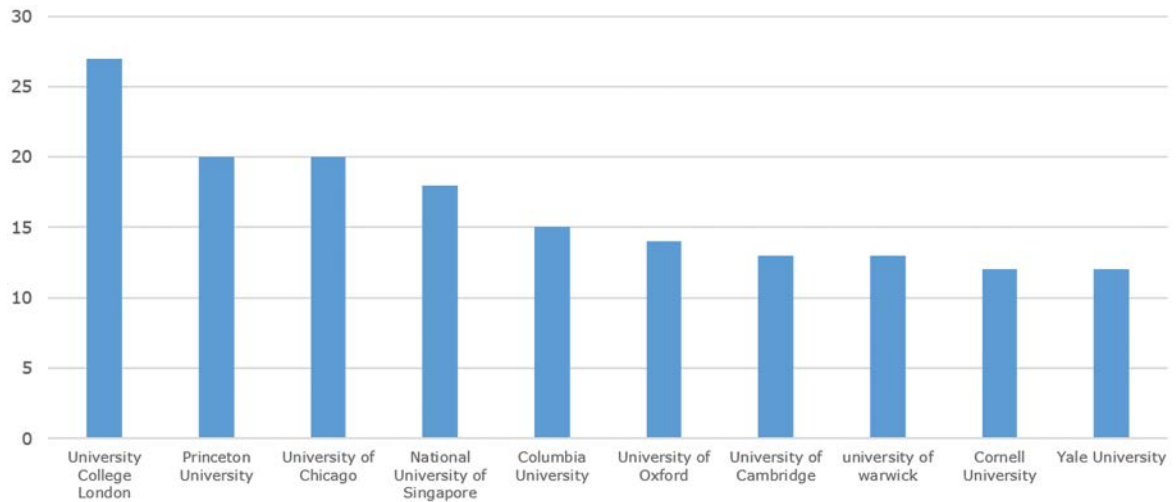**Figure 3.** Top 10 domestic institutions in terms of registered users.

Figure 4. Top 10 overseas institutions in terms of registered users.

3.4.2 Data Curation and Skills Training

To cultivate past research data for future consumption, the PKUL offered several data curation services. Collaborating with ISSS, the PKUL hosted an RDM Seminar in 2015. The PKUL invited two experts, one from the Inter-University Consortium for Political and Social Research (ICPSR), USA, and one from Data Archive, UK, respectively to deliver data management training. The trainees were teachers and students

from PKU, as well as from other peer universities in China. The PKUL also sent librarians to participate in relevant training activities hosted by other universities to improve their RDM curation skills. To improve students' data search skills, the PKUL offered one-hour workshops to teach students to identify data sources and use scientific methods to acquire relevant research data, statistical data, and Internet data. The PKUL also provided a series of lectures to teach students how to use data analytics tools.

3.4.3 Service Promotion

To improve the visibility of data service provided by the PKUL, the PKUL promoted the use of the PKU-ORDR through various channels, including marketing the PKU-ORDR on the PKU homepage, social media's public account of student groups, annual conferences of ISSS, and various RDM related domestic and international conferences. To improve data accessibility, the PKU-ORDR provided metadata to the re3data.org which is an international data repository registration, DataCite Search, and Data Citation Index which are data discovering systems, and search engines such as Baidu and Google. Additionally, collaborating with other units on campus and the National Information Center, and the PKUL hosted two national contests entitled "National Data-Driven Research Contests for Colleges and Universities" successively in the year 2018 and 2019 to promote the PKU-ORDR use and RDM services and train students' data searching and acquiring skills.

The contest included six stages: training workshops and lectures, enrollment, paper submission, paper evaluation, and oral defense. During the first stage, the organizers provided training on contest rules, data analysis and mining, data management and sharing, data resource, and acquisition. During the enrollment stage, the contestants registered in groups, submitted their selected topics, and applied for the research data in the PKU-ORDR. During the research paper submission stage, the contestants conducted research using the data from the PKU-ORDR or collected original data on their own, wrote essays, and submitted their papers together with the data to the organizers. During the paper evaluation stage, the organizers first conducted formal assessment and plagiarism checks for the essays, and the qualified papers then were evaluated by the experts invited by the organizers. Each essay was reviewed and graded by two experts. The papers were ranked accordingly by grades. During the stage of the on-spot oral defense, several top-ranked teams delivered their on-spot statements and reports and then were evaluated by more than 10 experts to decide the final ranking. After the contest, the winning teams shared their research data and reported at the Jing Ling Big Data Summits, and excellent essays were published in Chinese core journals in a special topic issue.

The contests attracted numerous students from many other major research universities in the country to participate. The first contest recorded an enrollment of nearly 600 teams including about 2,000 contestants from more than 160 universities and colleges. They came from 28 provinces, majoring in 59 disciplines such as Computer Science, Information Technology, Management Science & Engineering, Applied Economics, Statistics, Public Health & Preventive Medicine, Library and Information Science, and Chinese History. In the end, 289 teams including about 1,000 participants submitted their research papers. The second contest had an enrollment of 600 teams including 1,704 participants from 29 provinces, their

disciplines covering Applied Economics, Computer Science, Information Technology, Statistics, Sociology, Library and Information Science, Management Science & Engineering, Public Health & Preventive Medicine.

Through the contests, students from different disciplines obtained experiences from the same OD platform. The competition also greatly promoted the data-driven research paradigms and the visibility of the PKU-ORDR. Between December 2017 and May 2018 during the two contests, online visitors to the PKU-ORDR increased 10 times than before, registered users increased 5 times than before, and data downloads increased 7 times than before, respectively. More and more external websites were linked to PKU-ORDR. The ranking and exposure of the data in the PKU-ORDR are greatly improved in search engines. At the same time, the original data submitted by the contestants greatly enriched the content of the data repository.

4. CONCLUSIONS

This paper reviewed the implementation process of the PKU-ORDR and the creation of the RDM services provided by the PKUL. Through the review, the authors of the paper found that needs assessment and collaboration is vital to the success of a library-based university-wide RDM project. Raising the researchers' awareness to OS and OD is critical. Software identification and selection is a complicated and time-consuming process. The software must meet some essential criteria such as stability and sustainability. Communication is critical in the whole process, particularly with administrative units and other academic units on campus. Some data curation programs such as workshops, lectures, and contests can be developed to improve researchers', students', and librarians' data searching and acquiring skills and promote services on campus and to larger research communities. RDM policies must be created and put in place. In a word, this is a learning curve and a cumulative process in theories and practices. The authors of the paper will feel rewarded if this practical paper can offer some insights to those academic libraries planning to implement their OD repository and/or RDM services for their home institutions. Although the PKUL has made great efforts in the RDM construction and contributed to the OS and OD communities on campus and even in China, the PKUL feels that it still has a long way to go. There are so many challenges and opportunities ahead of libraries and librarians.

chinaXiv:202211.00398v1

and Conclusions, co-contributed to the layout, composition, and writing of the paper, and co-contributed to the utilization of sources.

REFERENCES

[1]   Burgelman, J.-C., et al.: Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. Frontiers in Big Data 2, Article 43 (2019)

[2]   Giglia, E.: Open Access to research data as a driver for Open Science. JLIS.it: Italian Journal of Library and Information Science 6(2), 225–247 (2015)

[3]   Tenopir, C., et al.: Changes in data sharing and data reuse practices and perceptions among scientists worldwide. PLoS ONE 10(8), e0134826 (2015)

[4]   Shen, Y. (2016). Research data sharing and reuse practices of academic faculty researchers: A study of the Virginia Tech data landscape. International Journal of Digital Curation 10(2), 157–175 (2016)

[5]   Schober, D., et al.: NmrML: A community supported open data standard for the description, storage, and exchange of NMR data. Analytical Chemistry 90(1), 649–656 (2018)

[6]   Tenopir, C., et al.: Research data management services in academic research libraries and perceptions of librarians. Library & Information Science Research 36(2), 84–90 (2014)

[7]   Cox, A. M., Pinfield, S.: Research data management and libraries: Current activities and future priorities. Journal of Librarianship and Information Science 46(4), 299–316 (2014)

[8]   Kruse, F., & Thestrup, J. B.: Research libraries' new role in research data management, current trends and visions in Denmark. LIBER Quarterly 23(4), 310–335 (2014)

[9]   Chiware, E., Mathe, Z.: Academic libraries' role in research data management services: A South African perspective. South African Journal of Library and Information Science 81(2), 1–10 (2016)

[10] Cox, A.M., et al.: Developments in research data management in academic libraries: Towards an understanding of research data service maturity. Journal of the Association for Information Science and Technology 68(9), 2182–2200 (2017)

[11] Ayris, P., Ignat, T.: Defining the role of libraries in the Open Science landscape: A reflection on current European practice. Open Information Science 2(1), 1–22 (2018)

[12] Hamad, F., Al-Fadel, M., Al-Soub, A.: Awareness of research data management services at academic libraries in Jordan: Roles, responsibilities and challenges. The New Review of Academic Librarianship, 1–21 (2019). Available at: https://doi.org/10.1080/13614533.2019.1691027. Accessed 22 December, 2020

[13] Pampel, H., et al.: Making research data repositories visible: The re3data.org registry. PLoS ONE 8(11), e78080 (2013)

[14] Moon, J.: Developing a research data management service – a case study. Partnership 9(1), 1–14 (2014)

[15] Curdt, C.: Supporting the interdisciplinary, long-term research project 'patterns in soil-vegetation-atmosphere-systems' by data management services. Data Science Journal 18(5), 1–9 (2019)

[16] Tripathi, M., Shukla, A., Sonkar, S. K.: Research data management practices in university libraries: A study. DESIDOC Journal of Library & Information Technology 37(6), 417 (2017)

[17] Johnston, L., et al.: Data curation network: How do we compare? A snapshot of six academic library institutions' data repository and curation services. Journal of Escience Librarianship 6(1), e1102 (2017)

[18] Lee, D.J., Stvilia, B.: Practices of research data curation in institutional repositories: A qualitative view from repository staff. PLoS ONE 12(3), e0173987 (2017)

[19] Curdt, C., Hoffmeister, D.: Research data management services for a multidisciplinary, collaborative research project. Program: Electronic Library and Information Systems 49(4), 494–512 (2015)

[20] McKinney, B., et al.: Extension of research data repository system to support direct compute access to biomedical datasets: Enhancing Dataverse to support large datasets. Annals of the New York Academy of Sciences 1387(1), 95–104 (2017)

[21] Mannheimer, S., et al.: Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges. The American Behavioral Scientist (Beverly Hills) 63(5), 643–664 (2018)

[22] Dovidonytė, R.: Implementation of Open Science in Lithuania. Nordic perspectives on Open Science, 1–12 (2019). Available at: https://doi.org/10.7557/11.4828. Accessed 22 December, 2020

[23] Pontika, N.: Roles and jobs in the open research scholarly communications environment: Analysing job descriptions to predict future trends. LIBER Quarterly 29(1), 1–20 (2019)

[24] Clare, C., et al.: Engaging researchers with data management: The cookbook. Open Book Publishers, Cambridge (2019). Available at https://doi.org/10.11647/OBP.0185.

[25] Alonso-Arévalo, J.: Research data management (RDM) on the horizon of academics and research libraries. Cuadernos De Documentación Multimedia 30, 75–88 (2019)

[26] Söderholm, M., Sunikka, A.: Collaboration in RDM activities – Practices and development at Aalto University. In: The 12th Munin Conference on Scholarly Publishing, pp. 1–10 (2017). Available at: https://doi.org/10.7557/5.4247. Accessed 22 December, 2020

[27] Tang, R., Hu, Z.: Providing research data management (RDM) services in libraries: Preparedness, roles, challenges, and training for RDM practice. Data and Information Management 3(2), 84–101 (2019)

[28] Zhu, L., et al.: (2016). The construction of Peking University open research data platform: Exploration and practice. Library and Information Service 60(4), 44–51 (2016) (in Chinese)

AUTHOR BIOGRAPHY

Hua Nie is Professor of Library Services at Peking University Library, China. Her research interests fall into the areas of open access, scholarly communication, digital humanities, and library marketing.
ORCID: 0000-0002-4522-5049

Pengcheng Luo is Data Librarian at Peking University Library, China. His research interests include data management and open data.
ORCID: 0000-0001-9598-0715
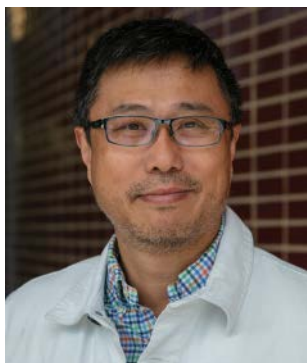
Ping Fu is Professor of Library Services and Science & Data Management Librarian at Central Washington University, USA. His research interests include integrated library systems, scholarly communication, and research data management.
ORCID: 0000-0002-5386-748X